arXiv:1602.00192v2 [cs.DS] 2 Feb 2016

On the Factor Revealing LP Approach for Facility Location with Penalties

Xian Qiu[1] *, Walter Kern[2]

[1] College of Computer Science, Zhejiang University
[2] Department of Applied Mathematics, University of Twente
xianqiu@zju.edu.cn, w.kern@utwente.nl

Abstract. We consider the uncapacitated facility location problem with (linear) penalty function and show that a modified JMS algorithm, combined with a randomized LP rounding technique due to Byrka-Aardal [1], Li [14] and Li *et al.* [16] yields 1.488 approximation, improving the factor 1.5148 due to Li *et al.* [16]. This closes the current gap between the classical facility location problem and this penalized variant. Main ingredient is a straightforward adaptation of the JMS algorithm to the penalty setting plus a consistent use of the upper bounding technique for factor revealing LPs due to Fernandes *et al.* [7]. In contrast to the bounds in [12], our factor revealing LP is monotone.

1 Introduction

An instance of the uncapacitated facility location problem is defined by a set C of *cities* and a set F of *facilities*. In addition, we are given *opening costs* f_i for each facility $i \in F$ as well as *connection costs* c_{ij} between facility $i \in F$ and city $j \in C$. The task is to open a subset of the facilities and to assign each city j to exactly one of the open facilities i (at a cost c_{ij}) in such a way that the total cost, *i.e.*, the total opening cost plus the total connection cost is minimized. We assume that the problem is *metric*, *i.e.*, the connection costs satisfy the triangle inequality $c_{ij} \leq c_{ij'} + c_{i'j'} + c_{i'j}$ for all facilities i, i' and cities j, j'.

A more general problem variant allows to *reject* certain cities j, *i.e.,* leave them unconnected, at a cost of $p_j > 0$. The goal is to open a subset of facilities and assign a subset of cities to the open facilities such that the total cost, *i.e.*, opening cost + connection cost + penalty cost, is minimized.

The (classical) facility location problem is NP-hard and has received a lot of attention in the literature. For history and detailed progress of the problem, *cf.* [11,9,5,13,2,19,18]. Jain *et al.* [12] presented an approximation algorithm – the JMS algorithm – achieving a factor 1.61. In their paper, they formally proposed the concept of "factor revealing LP", which is a linear program that reveals an upper bound of an approximation algorithm. Thus, the approximation ratio

* This work is supported by the Nature Science Foundation of Zhejiang Province, No. LQ15A010001 and the Fundamental Research Funds for the Central Universities of China.

can be found with the help of a computer program (*cf.* Section 3, 4). Building upon Chudak's LP rounding algorithm [5], Byrka and Aardal [1] introduced a clustering technique and presented an optimal bifactor $(\gamma, 1+2e^{-\gamma})$ – algorithm, with $\gamma > 1$. Here, a *bifactor* (γ_f, γ_c)-algorithm means that there exists an optimal solution with opening cost F^* and connection cost C^*, such that the total cost of the solution found by the algorithm is at most $\gamma_f F^* + \gamma_c C^*$. It was shown that for $\gamma = 1.67736$ the algorithm, combined with JMS, results in a 1.5-approximation [1]. Later, Li [14] proved that if γ is randomly selected according to some distribution, the approximation ratio can (again, in combination with the JMS algorithm) be improved to 1.488, which is the best known result currently. As to lower bounds, Guha and Khuller [9] showed that there is no λ-approximation algorithm for the metric facility location problem with $\lambda < 1.463$ unless NP $\subseteq$ DTIME$(n^{\log \log n})$.

Since the penalty variant generalizes the classical problem, both the hardness result and the inapproximability bound prevail to the penalty setting. The results on approximation can be summarized as follows. Charikar *et al.* [3] presented a primal-dual 3-approximation algorithm. Xu and Xu [20] improved the bound to 2.736 by LP-rounding. Later, Xu and Xu [21] developed a 1.853-approximation algorithm based on primal-dual and local search. Geunes *et al.* [8] described an approximation framework that converts any LP-rounding based α-approximation algorithm to a $(1-e^{-1/\alpha})^{-1}$ approximation algorithm. Building upon the work of Li [14], Li *et al.* [16] presented a 1.5148 approximation algorithm, which is currently best.

Some researchers also consider the problem with submodular penalties, where the total penalty is a submodular function on $2^{\mathcal{C}}$. The currently best bound is due to Li *et al.* [16], achieving a 2-approximation, via a combination of LP-rounding with the JMS algorithm. For other related results on submodular penalties, we refer to [10,4,6,15].

Our Contribution. We adapt the JMS algorithm to the penalty setting in the most straightforward way: The budget of each city j is raised only until it reaches p_j. But note that a rejected city may get connected at some time later. As it turns out, the corresponding factor revealing LPs are closely related to those in [12] and - at least numerically - yield similar bounds. In contrast to the factor revealing LPs in [12], however, ours are obviously monotone. Following the approach of Li [14], we combine the adapted JMS with Byrka and Aardal's clustering approach, which has been extended to the penalty setting by Li *et al.* [16], to achieve an overall bound of 1.488. In both cases (JMS and clustering) we apply the technique of Fernandez *et al.* [7] to the factor revealing LPs. In contrast to [7], however, we follow a purely "primal" approach, resulting in a more direct (and hopefully more intuitively appealing) derivation. Compared to the proofs in [14,16], this approach simplifies of the upper bound computation significantly.

The remaining part of the paper is organized as follows. In Section 2, we adapt the JMS algorithm to the penalty setting. The corresponding factor revealing

LP and the upper bound analysis are presented in Sections 3 and 4, respectively. In Section 5, we show that a combination of the modified JMS algorithm with randomized LP rounding as in [16,1,14] yields factor 1.488.

2 The Modified JMS Algorithm $\mathbb{A}_1$

The idea of the JMS algorithm can be interpreted as a cost sharing mechanism for the cities. In the classical version (without penalties) it works as follows. For each city j we determine a budget α_j, which is j's contribution for getting connected. Initially, all budgets are 0, all cities are unconnected and all facilities are closed. The algorithm proceeds by raising uniformly all budgets of cities that are still unconnected. So at time t, the budget of each unconnected city j is $\alpha_j = t$. When a city gets connected (as described below), its budget remains fixed from that point on.

Cities get connected in two possible ways: First, city j gets connected as soon as some open facility is "within reach", *i.e.,* is at distance $\leq \alpha_j$. Secondly, facilities get connected when they commonly open a facility. This works as follows. At any point in time, cities (connected or unconnected) offer certain amounts to closed facilities: Consider some closed facility i. At time t, each unconnected city j has a current budget of $\alpha_j = t$ and *offers* an amount of $[\alpha_j - c_{ij}]_+$ to facility i (where $[x]_+ = \max\{x, 0\}$). Each connected city j which is currently connected to some open facility at distance r_j offers an amount of $[r_j - c_{ij}]_+$ to facility i (the amount that it would save when switching to i). A closed facility i is opened as soon as it gets *fully paid*, *i.e.,* when the total offer (from both unconnected and connected) cities reaches its opening cost f_i. In this case, all cities that are offering a positive amount to i get connected to i. (Those that were already connected before switching to i.) Their offer to i is henceforth called their *contribution* toward the opening cost of i and cannot be withdrawn later on. Note that, as the total opening plus connection costs are fully paid from the budgets α_j, the solution constructed by the JMS algorithm has cost equal to $\sum_j \alpha_j$.

It is fairly obvious how to adapt this algorithm to the version with penalties: We simply raise each α_j only until either j gets connected or p_j is reached. Note that in the latter case, city j may still get connected at a later point in time ($t > \alpha_j$) in case a facility at distance $\leq \alpha_j$ gets opened. Hence, in contrast to the classical case (without penalties), α_j does not necessarily specify the point in time when city j gets connected for the first time. Indeed, in case the penalty p_j is relatively small, α_j stops raising quite early at time $t = p_j$. Yet city j keeps offering positive amounts to closed facilities at distance $< \alpha_j$ and some of them might be opened at a later point in time. Apart from this, however, there are no significant differences to the classical case. In particular, we like to note that according to the rules of the algorithm, no facility is ever "overpaid", *i.e.,* the total offer to a facility never exceeds its opening cost. Indeed, as soon as the total amount offered equals the opening cost, the facility is opened, the contributing cities get connected and their budgets are not raised any further.

3 The Factor Revealing LP

The approximation ratio of the adapted JMS algorithm can be estimated by a so-called *factor-revealing LP* [12]. The basic idea is to derive linear inequalities relating the problem data with certain parameters computed by the algorithm (such as the α_j) and then maximize the approximation ratio subject to these constraints.

More precisely, consider an optimal solution of our facility location problem. This can be described by specifying the corresponding set D of cities that are left unconnected and a collection $\mathcal{S}$ of *stars*, where each star $S \in \mathcal{S}$ consists of an open facility and all cities connected to it. Letting $\mathrm{cost}(S)$ denote the total cost of a star $S \in \mathcal{S}$ and $p(D) = \sum_{j \in D} p_j$ the total penalty cost of the optimal solution, the approximation ratio ρ of the adapted JMS algorithm is bounded by

$$\frac{\sum_{S \in \mathcal{S}} \sum_{j \in S} \alpha_j + \sum_{j \in D} \alpha_j}{\sum_{S \in \mathcal{S}} \mathrm{cost}(S) + p(D)} \leq \frac{\sum_S \sum_{j \in S} \alpha_j + p(D)}{\sum_S \mathrm{cost}(S) + p(D)} \leq \max_{S \in \mathcal{S}} \frac{\sum_{j \in \mathcal{S}} \alpha_j}{\mathrm{cost}(S)}. \quad (1)$$

Hence we may focus on a single star $S \in \mathcal{S}$ in the optimal solution, consisting of a facility with opening cost , say, f and the cities connected to it. In the following we will also denote this facilty with f (if no misunderstanding is possible) and use i, j etc. to denote cities. Let us assume that S contains exactly $k + l$ cities, say, cities $1, \ldots, k + l$ and let $d_j, j = 1, \ldots k + l$ denote their distances from f. Furthermore, assume that the adapted JMS algorithm connects cities $1, \ldots, k$ and leaves cities $k+1, \ldots, k+l$ unconnected. In addition, let us assume that cities $1, \ldots, k$ get connected in this order (ties broken arbitrarily).

Remark. In contrast to the classical setting (without penalties) this does not imply $\alpha_1 \leq \cdots \leq \alpha_k$. Indeed , a city j with small α_j may well get connected last!

In view of (1), we seek to bound $(\sum_{j=1}^{k+l} \alpha_j)/(f + \sum_{j=1}^{j=k+l} d_j)$, subject to some constraints that we derive next. Similarly, as observed in [12], if we are heading at a corresponding bifactor (γ_f, γ_c), we may consider a total of $\gamma_f f$ of the total budget as being spent for opening f and ask how large the resulting γ_c may become. In the generalized model with penalties, we combine the connection cost with the penalty cost. So a bifactor (γ_f, γ_c) means that the total cost is bounded by $\gamma_f F^* + \gamma_c(C^* + P^*)$. Thus, for fixed $\gamma_f \geq 1$, we would upper bound γ_c by $\max(\sum_j \alpha_j - \gamma_f f)/(\sum_j d_j)$, subject to the constraints derived below.

To derive the constraints, consider any $j < i \leq k$. Thus city j is already connected to some open facility at distance, say, r_{ji} immediately before city i gets connected (for the first time). Clearly, when j gets connected for the first time, it connects to an open facility at distance at most α_j. Thus we have $\alpha_j \geq r_{j,j+1}$. Afterwards, j may switch only to closer facilities, so we know that

$$\alpha_j \geq r_{j,j+1} \geq r_{j,j+2} \geq \ldots \geq r_{j,k}. \tag{2}$$

Next, we use the fact mentioned at the end of section 2 that facility f does not get overpaid. Thus, in particular, the total offer from cities $1, \ldots, k+l$ does not exceed the opening cost f. At the point in time immediately before i gets connected, each city $j \in \{i, i+1, \ldots k+l\}$ is still unconnected and hence offers an amount of $[\alpha_j - d_j]_+$ to f. Each city $j \in \{1, \ldots i-1\}$ is already connected and its current connection cost equals r_{ji}, so it offers an amount of $[r_{ji} - d_j]_+$. This gives rise to an *opening cost constraint*

$$\sum_{j=1}^{i-1}[r_{j,i} - d_j]_+ + \sum_{j=i}^{k+l}[\alpha_i - d_j]_+ \leq f, \quad i = 1, \ldots, k \tag{3}$$

Finally, there are upper bounds due to the triangle inequality:

$$\alpha_i \leq r_{ji} + d_j + d_i, \quad 1 \leq j < i \leq k. \tag{4}$$

Indeed, note that immediately before i gets connected, j is already connected to some open facility at distance r_{ji}. The distance from i to this open facility is at most $d_i + d_j + r_{ji}$, hence α_i will never be raised above this value.

Now we are ready to formulate the bifactor revealing LP (for fixed $\gamma_f \geq 1$).

$$\begin{aligned}
\gamma_c = \quad & \max\ (\sum_{i=1}^{k+l} \alpha_i - \gamma_f f)/(\sum_{i=1}^{k+l} d_i) \\
\text{s.t.}\quad & \alpha_j \geq r_{j,j+1} \geq \cdots \geq r_{jk} \quad \forall\ 1 \leq j < k \\
& \alpha_i \leq r_{j,i} + d_i + d_j, \quad \forall 1 \leq j < i \leq k \\
& \sum_{j=1}^{i-1} [r_{j,i} - d_j]_+ + \sum_{j=i}^{k+l}[\alpha_j - d_j]_+ \leq f, \quad \forall\ 1 \leq i \leq k \\
& \alpha_j, d_j, f, r_{j,i} \geq 0, \quad \forall\ 1 \leq j \leq i \leq k+l.
\end{aligned} \tag{5}$$

In what follows we will show that the maximum is achieved when there are no cities $j \in k+1, \ldots k+l$, *i.e.,* when JMS actually connects all cities in the star. Indeed, consider an optimal solution (α, d, f, r) and a fixed $j \in \{k+1, \ldots, k+l\}$. In case $\alpha_j \leq d_j$ we may decrease d_j to α_j, thereby increasing the objective value (as the denominator decreases). Note that none of the constraints will get violated as the triangle constraints only involve d_j's with $j \leq k$. Hence we may assume $\alpha_j \geq d_j$. In this case we decrease both α_j and d_j by the same amount, yielding again a higher objective value, unless actually $\alpha_j \geq d_j = 0$ holds. In case $\alpha_j > 0$ we may decrease both α_j and f by the the same amount, again leading to an increase in the objective. This shows that also $\alpha_j = 0$ can be assumed and hence city j is completely redundant. So we may indeed simply remove the cities $k+1,\dots,k+l$ from our considerations.

Summarizing, (5) is (with $\sum_j d_j$ scaled to 1) equivalent to

$$\begin{aligned} \gamma_c^k = \quad & \max \sum_{i=1}^{k} \alpha_i - \gamma_f f \\ & \text{s.t. } \sum_{i=1}^{k} d_i = 1, \\ & \alpha_j \geq r_{j,j+1} \geq \cdots \geq r_{jk} \quad \forall 1 \leq j < k \\ & \alpha_i \leq r_{j,i} + d_i + d_j, \quad \forall 1 \leq j < i \leq k \\ & \sum_{j=1}^{i-1} [r_{j,i} - d_j]_+ + \sum_{j=i}^{k} [\alpha_i - d_j]_+ \leq f, \quad \forall 1 \leq i \leq k \\ & \alpha_j, d_j, f, r_{j,i} \geq 0, \quad \forall 1 \leq j \leq i \leq k. \end{aligned} \tag{6}$$

Remark. The above LP differs from the one in [12] in that we do not require $\alpha_1 \leq \cdots \leq \alpha_k$. Instead, we included the constraints $\alpha_j \geq r_{j,j+1}$ which do not appear in [12]. (In [12] they are claimed to be redundant.)

A remarkable difference to [12] is that our upper bounds γ_c^k are monotonically increasing in k: Setting $\alpha_{k+1} = d_{k+1} = r_{j,k+1} = 0$ for all $j = 1,\dots,k$, we can extend any optimal solution of (6) to a feasible solution with $k+1$ cities. Hence indeed, $\gamma_c^k \leq \gamma_c^{k+1}$ must hold. As we will see, the γ_c^k are bounded, so $\lim_{k\to\infty} \gamma_c^k$ exists and hence JMS has a bifactor (γ_f, γ_c) with $\gamma_c \leq \lim_{k\to\infty} \gamma_c^k$.

4 The Upper Bound

In the original analysis of Jain *et al.* [12], the upper bound on a factor-revealing LP is analyzed by using suitable "dual parameters", which are estimated by computational experiments, to find a feasible solution to the dual program for any large k. This approach was later improved by Fernandes *et al.* [7], deriving an *upper bound factor-revealing LP*, *i.e.*, an LP whose optimal value is an upper bound of every factor-revealing LP. We follow their idea, but present a corresponding "primal approach" based on "aggregating variables". (As we found out later, a similar approach was developed by Mahdian and Yan [17], even before Fernandes *et al.* [7]) in another context.)

Intuitively, the reason why γ_c^k grows rather slowly is that among a large number k of cities, there must be many with the same characteristics (*i.e.*, distances from facility locations). Aggregating cities with almost the same characteristics decreases k considerably without affecting γ_c^k a lot. Formally, this corresponds to *aggregating variables* as in Theorem 1 below. This is, roughly, the "primal" idea behind the approach presented in [7].

Theorem 1. *Let $\hat{k}$ be a positive integer and $\gamma_f \geq 0$. Then*

$$\begin{aligned}
\hat{\gamma}_c \quad = \quad & \max \sum_{t=1}^{\hat{k}} \alpha_t - \gamma_f f \qquad\qquad (7)\\
s.t. \quad & \sum_{s=1}^{\hat{k}} d_s = 1\\
& \alpha_s \geq r_{s,s+1} \geq \cdots \geq r_{s,\hat{k}} \quad \forall 1 \leq s < \hat{k}\\
& \alpha_s \leq r_{t,s} + d_s + d_t, \quad \forall 1 \leq t < s \leq \hat{k}\\
& \sum_{t=1}^{s-1} [r_{t,s} - d_t]_+ + \sum_{t=s+1}^{\hat{k}} [\alpha_s - d_t]_+ \leq f, \quad \forall 1 \leq s \leq \hat{k}\\
& \alpha_s, d_s, f, r_{t,s} \geq 0, \quad \forall 1 \leq t \leq s \leq \hat{k}
\end{aligned}$$

provides an upper bound on γ_c (corresponding to the prescribed γ_f).

Remark: At the first glance, one may wonder whether there is any difference between (6) and (7) at all. Indeed, (7) is only a slight relaxation of (6), obtained by dropping the term $[\alpha_s - d_t]_+$ for $t = s$ in the opening cost constraints.

Proof. As γ_c^k is increasing, it suffices to show $\hat{\gamma}_c \geq \gamma_c^{\hat{k}p}$ for all $p \in \mathbb{N}$. Let f, α_i, d_i, $r_{j,i}$, $1 \leq j \leq i \leq \hat{k}p$ be an optimal solution of (6) for $k = \hat{k}p$. Note that variables $r_{j,i}$ are defined with indices $1 \leq j < i \leq \hat{k}p$. For variables indexed in $1, \ldots, \hat{k}p$, we *aggregate* sets of p variables into one by setting

$$\begin{aligned}
& I_s := \{(s-1)p + 1, \ldots, sp\}, \ s = 1, \ldots, \hat{k},\\
& \hat{\alpha}_s := \sum_{i \in I_s} \alpha_i, \ \hat{d}_s := \sum_{i \in I_s} d_i, \ \hat{r}_{t,s} := \frac{1}{p} \sum_{j \in I_t, i \in I_s} r_{j,i}, \ 1 \leq t < s \leq \hat{k}.
\end{aligned}$$

We aim at showing that $(f, \hat{\alpha}_s, \hat{d}_s, \hat{r}_{t,s})$ is feasible for (7) and its objective value is at least γ_c^k. First note that $\sum_{j=1}^{k} \alpha_j = \sum_{t=1}^{\hat{k}} \hat{\alpha}_t, \quad \sum_{j=1}^{k} d_j = \sum_{t=1}^{\hat{k}} \hat{d}_t$. Hence, in particular , the objective function in (6) and (7) have the same value. Thus it suffices to show that $(f, \hat{\alpha}_s,\ \hat{d}_s,\ \hat{r}_{t,s})$ is feasible for (7).

This is straightforward to verify: The first two constraints are obviously satisfied. As to the third, observe that for $j \in I_s$ and $i \in I_{s+1}$, $\alpha_j \geq r_{j,i}$. Thus $\alpha_j \geq \frac{1}{p} \sum_{i \in I_{s+1}} r_{j,i}$ and hence $\sum_{j \in I_s} \alpha_j \geq \frac{1}{p} \sum_{j \in I_s} \sum_{i \in I_{s+1}} r_{j,i}$, *i.e*, $\hat{\alpha}_s \geq \hat{r}_{s,s+1}$ indeed.

As to the triangle inequality constraints, note that for $1 \leq t < s \leq \hat{k}$,

$$\begin{aligned}
p\hat{\alpha}_s &= \sum_{j=1}^{p} \sum_{i=1}^{p} \alpha_{(s-1)p+i}\\
&\leq \sum_{j=1}^{p} \sum_{i=1}^{p} r_{(t-1)p+j,(s-1)p+i} + \sum_{j=1}^{p} \sum_{i=1}^{p} d_{(s-1)p+j} + \sum_{j=1}^{p} \sum_{i=1}^{p} d_{(t-1)p+i}
\end{aligned}$$

$$= p\hat{r}_{t,s} + p\hat{d}_s + p\hat{d}_t,$$

yielding the desired inequality.

Finally, consider the opening cost constraints: Summing up the original opening cost constraints in (6) for $i = (s-1)p+1, \ldots, sp$ yields

$$\begin{aligned}
pf &\geq \sum_{i=(s-1)p+1}^{sp} \sum_{j=1}^{i-1} [r_{j,i} - d_j]_+ + \sum_{i=(s-1)p+1}^{sp} \sum_{j=i}^{\hat{k}p} [\alpha_i - d_j]_+ \\
&\geq \sum_{i=(s-1)p+1}^{sp} \sum_{j=1}^{(s-1)p} [r_{j,i} - d_j]_+ + \sum_{i=(s-1)p+1}^{sp} \sum_{j=sp+1}^{\hat{k}p} [\alpha_i - d_j]_+ \\
&\geq \sum_{t=1}^{s-1} \sum_{j=1}^{p} \sum_{i=1}^{p} [r_{(t-1)p+j,(s-1)p+i} - d_{(t-1)p+j}]_+ \\
&\quad + \sum_{t=s+1}^{\hat{k}} \sum_{j=1}^{p} \sum_{i=1}^{p} [\alpha_{(s-1)p+i} - d_{(t-1)p+j}]_+ \\
&\geq p \cdot \sum_{t=1}^{s-1} [\hat{r}_{t,s} - \hat{d}_t]_+ + p \cdot \sum_{t=s+1}^{\hat{k}} [\hat{\alpha}_s - \hat{d}_t]_+,
\end{aligned}$$

where the last inequality holds as $[a]_+ + [b]_+ \geq [a+b]_+$. Hence, indeed, the opening cost constraints in (7) are satisfied as well.

Solving the above LP for $k = 500$ and $\gamma_f = 1.107$, we obtain a bifactor $(1.107, 1.780)$.

5 Randomized LP Rounding $\mathbb{A}_2(\gamma)$

Byrka and Aardal [1] proposed an LP-rounding technique that starts from the canonical LP relaxation of the classical facility location problem. Later, this approach was extended in Li *et al.* [16] to include penalties. The corresponding LP looks as follows (z_j denotes the fraction of city j that is rejected):

$$\begin{aligned}
\min\ & \sum_{i\in\mathcal{F}} \sum_{j\in\mathcal{C}} c_{ij} x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i + \sum_{j\in\mathcal{C}} p_j z_j \\
\text{s.t.}\ & \sum_{i\in\mathcal{F}} x_{ij} + z_j = 1, \quad \forall j \in \mathcal{C} \\
& x_{ij} \leq y_i, \quad \forall i \in \mathcal{F},\ \forall j \in \mathcal{C} \\
& x_{ij}, y_i, z_j \geq 0, \quad \forall i \in \mathcal{F},\ \forall j \in \mathcal{C}.
\end{aligned} \tag{8}$$

Note that the crucial variables in (8) are the y-variables. Indeed, for given y, the best corresponding x, z are obtained by fractionally connecting each city j

to its closest fractionally open facilities, up to a distance of p_j. The fraction of j that is left unconnected is z_j.

The basic idea in [1] and [16] is to first compute an optimal solution (y^*, x^*, z^*) of (8), then replace y^* by $y = \gamma y^*$ for some $\gamma \geq 1$, determine the corresponding x, z variables and "round" this solution to an integral one. One may assume that at least one city is fully connected even in the original solution (y^*, x^*, z^*). Otherwise, if $z_j^* \geq z_{\min} > 0$ for all j, we would design a randomized algorithm that with probability $z_{\min}$ rejects all cities, and, with probability $1 - z_{\min}$, applies rounding as described below to the solution defined by $\frac{1}{1-z_{\min}} y^*$. (Actually, this observation is missing in Li *et al.* [16].) The rounding is based on y: For each city j define its "neighbourhood" $F_j = \{i \in F | x_{ij} > 0\}$. Among the cities that are fully connected in the solution (y, x, z), greedily choose a set of "centers" with pairwise disjoint F_j's. Then open facilities independently at random with probability y_i within the neighbourhood of each center. Thus, within each center neighbourhood at least one facility is opened. Open remaining facilities outside these neighbourhoods independently at random with probability y_i. This defines $0-1$ values for the y-variables and the corresponding optimal choice for the x, z variables will be $0-1$ automatically. Note that the expected opening cost equals $\sum_i f_i y_i = \gamma \sum_i f_i y_i^*$, *i.e.*, the algorithm has a bifactor (γ_f, γ_c) with $\gamma_f = \gamma$. Bounding the corresponding γ_c for given $\gamma_f = \gamma$ is fairly involved. The basic argument in [16] is identical to the reasoning in the classical case without penalties (*cf.* [14]): Introduce a fictive "rejection" facility at distance p_j from each city j. Interpret opening of such a facility as rejecting j. However, these facilities are not part of the metric space our instance lives in. In particular, the algorithm will never connect any other city except j to its rejection facility and, correspondingly, does not rely on any triangle inequalities involving rejection facilities. We only present the final outcome here.

Each city j has an associated "characteristic" function $h_j : [0,1] \to \mathbb{R}$ defined implicitly by the LP solution (y^*, x^*, z^*). The function h_j describes how the marginal connection cost increases with the fraction of j that we want to connect – given that facilities are opened according to y^*. (For $p \geq 1 - z_j^*$, define $h_j(p) = p_j$.) In other words, the function h_j (which is easily read off from the LP solution) is such that $\int_0^p h_j(u)du$ is the connection plus penalty cost for a p-fraction of city j in the LP-solution. The crucial result (*cf.* Lemma 4.3 in [16]) states that the expected connection plus penalty cost for city j (expectation taken *w.r.t.* random opening facilities with probabilities $y_i = \gamma y_i^*$ for fixed $\gamma \in [1,2]$) can be bouned by

$$\mathbb{E}(C_j + P_j) \leq \int_0^1 h_j(u) e^{-\gamma u} \gamma du + e^{-\gamma} \left[\gamma \int_0^1 h_j(u)du + (3-\gamma) h_j(1/\gamma) \right].$$

Letting $h = \sum_j h_j$, we conclude that $\int_0^1 h(u)du$ equals the connection plus penalty cost of the LP solution. Scaling $\int_0^1 h(u)du = 1$, we thus conclude that

γ_c is bounded by

$$\gamma_c \leq \ \alpha(\gamma, h) := \int_0^1 h(u)e^{-\gamma u}\gamma du + e^{-\gamma}\left[\gamma + (3-\gamma)h(1/\gamma)\right]. \tag{9}$$

Now h is a step function, that splits into single-step functions of the form

$$h_q(u) = \begin{cases} 0, & u \leq q \\ \frac{1}{1-q}, & u > q \end{cases}, \qquad \text{with } 0 \leq q < 1$$

and hence, by linearity, the largest upper bound in (9) is obtained by replacing h with one of these h_q.

Now that we have determined bifactors (a, b) for the JMS algorithm $\mathbb{A}_1$ and (γ, γ_c) for the LP-rounding algorithm $\mathbb{A}_2(\gamma)$, it is natural to combine these two by randomly choosing, say, $\mathbb{A}_1$ with some probability θ and $\mathbb{A}_2(\gamma)$ with probability $1-\theta$, as in [1]. Even better, as shown in [14], one should choose γ from a suitable distribution, say, μ on $[1, 2]$ with $\int_1^2 \mu(\gamma)d\gamma = 1-\theta$ and obtain

$$\mathbb{E}(\gamma_f) \leq \int_1^2 \gamma\mu(\gamma)d\gamma + a\theta, \qquad \mathbb{E}(\gamma_c) \leq \int_1^2 \alpha(\gamma, h_q)\mu(\gamma)d\gamma + b\theta. \tag{10}$$

We discretize the domain $[1, 2]$ of γ and let $\gamma_i = 1 + \frac{i}{n}$, for $i = 1, \ldots, n$. Also discretize the domain $[0, 1)$ of q by letting $q_j = j/m$ for $j = 0, \ldots, m-1$. At any point γ_i and q_j, we can compute

$$\alpha_{ij} := \alpha(\gamma_i, h_{q_j}) = \frac{1}{1-q_j}\left(e^{-\gamma_i q_j} - e^{-\gamma_i}\right) + e^{-\gamma_i}\left[\gamma_i + (3-\gamma_i)h_{q_j}(1/\gamma_i)\right].$$

By definitions of h_q, γ_i and q_j, the above inequality can be written as

$$\alpha_{ij} = \begin{cases} \frac{1}{1-q_j}\left(e^{-\gamma_i q_j} - e^{-\gamma_i}\right) + e^{-\gamma_i}\gamma_i, & \frac{1}{\gamma_i} \leq q \\ \frac{1}{1-q_j}\left(e^{-\gamma_i q_j} - e^{-\gamma_i}\right) + e^{-\gamma_i}\left[\gamma_i + (3-\gamma_i)\frac{1}{1-q_j}\right], & \frac{1}{\gamma_i} > q \end{cases}$$

For $i = 1, \ldots, n$ and $j = 1, \ldots, m-1$, define

$$u_{ij} = \frac{1}{1-q_j}\left(e^{-\gamma_i q_j} - e^{-\gamma_i}\right) + e^{-\gamma_i}\gamma_i \text{ and } v_{ij} = e^{-\gamma_i}(3-\gamma_i)\frac{1}{1-q_j}.$$

Thus, $\alpha_{ij} = u_{ij} + v_{ij}$ if $\gamma_i q_j < 1$ and $\alpha_{ij} = u_{ij}$ otherwise. Let $\mu_i := \mu(\gamma_i)$ for $i = 1, \ldots, n$. According to (10), the approximation ratio is approximated by

$$\min \beta \tag{11}$$

$$\text{s.t. } \frac{1}{n}\sum_{i=1}^{n} \mu_i = 1-\theta$$

$$\frac{1}{n}\sum_{i=1}^{n}\gamma_i\mu_i + a\theta \le \beta$$
$$\frac{1}{n}\sum_{i=1}^{n}u_{ij}\mu_i + \frac{1}{n}\sum_{i:\gamma_i q_j<1}v_{ij}\mu_i + b\theta \le \beta, \quad \forall j=0,\ldots,m-1$$
$$\mu_1,\ldots,\mu_n,\theta \ge 0.$$

Let $\beta_{n,m}$ be the optimal objective function value of the above LP. Thus $\lim_{n,m\to\infty}\beta_{n,m}$ equals the approximation ratio. Thus we are to find an upper bound for the factor-revealing LP (11). Again, this can be achieved by a slight modification of the original LP:

Theorem 2. *Define*

$$\hat{u}_{ij} = \frac{1}{1-q_j}\left(e^{-\gamma_{i-1}q_{j-1}} - e^{-\gamma_i}\right) + e^{-\gamma_{i-1}}\gamma_{i-1} \text{ and } \hat{v}_{ij} = e^{-\gamma_{i-1}}(3-\gamma_{i-1})\frac{1}{1-q_j}.$$

Then an upper bound on the approximation ratio is obtained by solving

$$\hat{\beta} = \min\ \beta \tag{12}$$
$$s.t.\ \frac{1}{n}\sum_{i=1}^{n}\mu_i = 1-\theta$$
$$\frac{1}{n}\sum_{i=1}^{n}\gamma_i\mu_i + a\theta \le \beta$$
$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_{ij}\mu_i + \frac{1}{n}\sum_{i:\gamma_{i-1}q_{j-1}<1}\hat{v}_{ij}\mu_i + \quad b\theta \le \beta, \quad \forall j=1,\ldots,m-1$$
$$\mu_1,\ldots,\mu_n,\theta \ge 0$$

Proof. First note that the constraints in (12) are stronger compared to (11): Monotonicity of the functions $\frac{1}{1-q}$ (increasing), $e^{-\gamma}$ and $e^{-\gamma}\gamma$ (decreasing), as well as $e^{-\gamma q}$ (decreasing in both γ and q) imply $\hat{u}_{ij} \ge u_{ij}$ and $\hat{v}_{ij} \ge v_{ij}$. Moreover, the condition $\gamma_{i-1}q_{j-1} < 1$ is weaker than $\gamma_i q_j < 1$, so in the last constraint in (12) the summation is over more terms than in (11). This proves that indeed the optimum value in (12) must be larger (for identical values of m, n!).

Next let us consider (12) for some fixed parameters $\hat{m}, \hat{n}$ and let $\hat{\mu}, \theta$ be a corresponding optimal solution with objective value $\hat{\beta}$. Define the corresponding discretized values of γ and q by $\hat{\gamma}_{\hat{i}} = 1 + \frac{\hat{i}}{\hat{n}},\ \hat{i} = 0,\ldots\hat{n}$ and $\hat{q}_{\hat{j}} = \frac{\hat{j}}{\hat{m}},\ \hat{j} = 0,\ldots,\hat{m}-1$. We claim that $\hat{\beta} \ge \beta_{n,m}$ for $n,m\to\infty$. It suffices to prove this inequality for $n = p\hat{n}, m = p\hat{m},\ p \ge 1$. Recall that $\gamma_i = 1 + \frac{i}{n}, i = 1,\ldots,n$ and $q_j = \frac{j}{m},\ j = 1,\ldots,m-1$. As in section 4 we define $I_{\hat{i}} = \{(\hat{i}-1)p+1,\ldots,\hat{i}p\}$ and $J_{\hat{j}} = \{(\hat{j}-1)p+1,\ldots,\hat{j}p\}$. Monotonicity of the functions $\frac{1}{1-q}$, $e^{-\gamma}$, $e^{-\gamma}\gamma$

and $e^{-\gamma q}$ then ensures that

$$\hat{u}_{\hat{i}\hat{j}} \geq u_{ij} \quad \text{and } \hat{v}_{\hat{i}\hat{j}} \geq v_{ij} \text{ for all } i \in I_{\hat{i}} \text{ and } j \in J_{\hat{j}}. \tag{13}$$

Hence defining $\mu_i := \hat{\mu}_{\hat{i}}$ for all $i \in I_{\hat{i}}$ yields a feasible solution μ, θ for (11) with

$$\frac{1}{n}\sum_i \mu_i = \frac{1}{\hat{n}}\sum_{\hat{i}} \hat{\mu}_{\hat{i}} = 1 - \theta,$$
$$\frac{1}{n}\sum_i \gamma_i \mu_i \leq \frac{1}{\hat{n}}\sum_{\hat{i}} \hat{\gamma}_{\hat{i}} \hat{\mu}_{\hat{i}}.$$

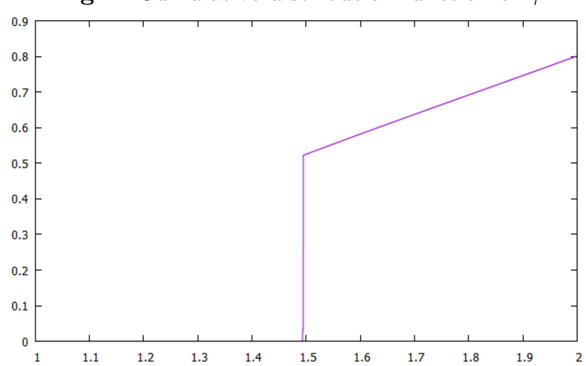
Finally, for $j \in J_{\hat{j}}$ we find (using (13))

$$\sum_{i=1}^{n} u_{ij}\mu_i + \sum_{i:\gamma_i q_j < 1} v_{ij}\mu_i \ \leq \ p \left[\sum_{\hat{i}=1}^{\hat{n}} \hat{u}_{\hat{i}\hat{j}}\hat{\mu}_{\hat{i}} + \sum_{\hat{i}:\hat{\gamma}_{\hat{i}-1}\hat{q}_{\hat{j}-1} < 1} \hat{v}_{\hat{i}\hat{j}}\hat{\mu}_{\hat{i}} \right].$$

Note that here we also use the fact that for $i \in I_{\hat{i}}$ and $j \in J_{\hat{j}}$ the inequality $\gamma_i q_j < 1$ implies $\hat{\gamma}_{\hat{i}-1}\hat{q}_{\hat{j}-1} < 1$. Thus the lower bounds for β in (12) are indeed stronger than those in (11), proving the claim.

Referring to the end of Section 4, we know that the modified JMS algorithm has a bifactor $(1.107, 1.780)$. Solving the above LP with $m = n = 8000$ yields an upper bound 1.48786, with $\theta \approx 0.19878$ and the distribution of γ (*cf.* Fig.1).

Fig. 1. Cumulative distribution function of γ



References

1. J. Byrka and K. Aardal. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. *SIAM Journal on Computing*, 39(6):2212–2231, 2010.

2. M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *Proceedings of FOCS' 99*, pages 378–388, 1999.
3. M. Charikar, S. Khuller, D. Mount, and G. Narasimhan. Algorithms for facility location problems with outliers. In *Proceedings of SODA' 01*, pages 642–651, 2001.
4. F. Chudak and K. Nagano. Efficient solutions to relaxations of combinatorial problems with submodular penalties via the lovász extension and non-smooth convex optimization. In *Proceedings of SODA' 07*, pages 79–88, 2007.
5. F. A. Chudak. Improved approximation algorithms for uncapacitated facility location. In *Proceedings of IPCO*, volume 1412 of *LNCS*, pages 180–194. 1998.
6. D. Du, R. Lu, and D. Xu. A primal-dual approximation algorithm for the facility location problem with submodular penalties. *Algorithmica*, 63:191–200, 2012.
7. C. G. Fernandes, L. A. A. Meira, F. K. Miyazawa, and L. L. C. Pedrosa. A systematic approach to bound factor revealing LPs and its application to the metric and squared metric facility location problems. In *Proceedings of APPROX/RANDOM' 12*, pages 146–157, 2012.
8. J. Geunes, R. Levi, H. Romeijn, and D. Shmoys. Approximation algorithms for supply chain planning and logistics problems with market choice. *Mathematical Programming*, 130:85–106, 2011.
9. S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, 31(1):228 – 248, 1999.
10. A. Hayrapetyan, C. Swamy, and E. Tard́os. Network design for information networks. In *Proceedings of SODA' 05*, pages 933–942, 2005.
11. D. Hochbaum. Heuristics for the fixed cost median problem. *Mathematical Programming*, 22:148–162, 1982.
12. K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *J. ACM*, 50(6):795–824, 2003.
13. K. Jain and V. V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *J. ACM*, 48(2):274–296, 2001.
14. S. Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45–58, 2013.
15. Y. Li, D. Du, N. Xiu, and D. Xu. A combinatorial 2.375-approximation algorithm for the facility location problems withsubmodular penalties. *Theoretical Computer Science*, 476:109–117, 2013.
16. Y. Li, D. Du, N. Xiu, and D. Xu. Improved approximation algorithms for the facility location problems with linear/submodular penalties. *Algorithmica*, pages 460–482, 2015.
17. M. Mahdian and Q. Yan. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing LPs. In *Proceedings of STOC' 11*, pages 597–606, 2011.
18. M. Mahdian, Y. Ye, and J. Zhang. Improved approximation algorithms for metric facility location problems. In *Proceedings of APPROX '02*, pages 229–242, 2002.
19. M. Sviridenko. An improved approximation algorithm for the metric uncapacitated facility location problem. In *Proceedings of IPCO*, volume 2337 of *LNCS*, pages 240–257. 2006.
20. G. Xu and J. Xu. An LP rounding algorithm for approximating uncapacitated facility location problem with penalties. *Information Processing Letters*, 94:119–123, 2005.

21. G. Xu and J. Xu. An improved approximation algorithm for uncapacitated facility location problem with penalties. *Journal of Combinatorial Optimization*, 17:424–436, 2008.